The genomic basis of circadian and circalunar timing adaptations in a midge

Tobias S. Kaiser[1,2,3,9], Birgit Poehn[1,3], David Szkiba[2], Marco Preussner[4], Fritz J. Sedlazeck[2,10], Alexander Zrim[2], Tobias Neumann[1,2], Lam-Tung Nguyen[2,5], Andrea J. Betancourt[6], Thomas Hummel[3,7], Heiko Vogel[8], Silke Dorner[1], Florian Heyd[4], Arndt von Haeseler[2,3,5], Kristin Tessmar-Raible[1,3]

[1] Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter, Dr. Bohr-Gasse 9/4, 1030 Vienna, Austria

[2] Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna and Medical University of Vienna, Dr. Bohr-Gasse 9, 1030 Vienna, Austria

[3] Research Platform ‘‘Rhythms of Life,’’ University of Vienna, 1030 Vienna, Austria

[4] Department of Biology, Chemistry, Pharmacy, Institute of Chemistry and Biochemistry, FU Berlin, 14195 Berlin, Germany

[5] Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

[6] Institute of Population Genetics, Department of Biomedical Sciences, University of Veterinary Medicine Vienna, Josef-Baumann-Gasse 1, 1210 Vienna, Austria

[7] Dept. of Neurobiology, Faculty of Life Sciences, University of Vienna, 1090 Vienna, Austria

[8] Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knöll-Straße 8, 07745 Jena, Germany

[9] current address: Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany

[10] current address: Department of Computer Science, Johns Hopkins University, Baltimore, MD 21211, USA

Contact Information

kristin.tessmar@mfpl.ac.at,

kaiser@evolbio.mpg.de

Good timing is crucial in life. Thus, organisms use endogenous clocks to anticipate regular environmental cycles, such as days and tides. Natural variants resulting in differently timed behavior or physiology, known as chronotypes in humans, are little understood at the molecular level. We generated a high-quality reference genome of *Clunio marinus*, a marine midge whose reproduction is timed by circadian and circalunar clocks. Midges from different locations show strain-specific genetic timing adaptations. By studying genomic sequence variations in five *C. marinus* strains we identified genes associates with circalunar and circadian chronotypes, none encoding core circadian clock genes. The region most strongly associated with circadian chronotypes generates strain-specific abundance differences of *Ca^{2+}/Calmodulin-dependent kinase II.1* (*CaMKII.1*) splice variants. As equivalent variants were shown to alter CaMKII activity in *D.melanogaster,* and *Cma*-CaMKII.1 increases the transcriptional activity of the *Cma*-Clock*/Cma*-Cycle, we suggest alternative splicing modulation as a mechanism for natural adaptation in circadian timing.

Around new or full moon, during a few specific hours surrounding low tide, millions of non-biting midges of the species *Clunio marinus* emerge out of the sea to perform their nuptial dance. Adults live only a few hours, during which they mate and oviposit. They must therefore emerge synchronously and – given that embryonic, larval and pupal development take place in the sea – at a time when the most extreme tides reliably expose the larval habitat. The lowest low tides occur predictably during specific days of the lunar month at a specific time of day. Consequently, adult emergence in *C. marinus* is under the control of circalunar and circadian clocks[1,2]. Importantly, while the lowest low tides recur

invariably at a given location, their timing differs between geographic locations[3]. Congruently, *C. marinus* strains from different locations (Extended Data Fig. 1a) show local adaptation in circadian and circalunar emergence times (Extended Data Fig. 1b,c). Crosses between the *Jean* and *Por* strains showed that the differences in circadian and circalunar timing are genetically determined[4,5] and largely explained by two circadian and two circalunar quantitative trait loci (QTLs)[6].

Studies on timing variation or chronotypes in animals and humans have often focused on candidate genes from the circadian transcription-translational oscillator: In *D. melanogaster*, polymorphisms in the core circadian clock genes *period, timeless* and *cryptochrome* are associated with adaptive differences in temperature compensation[7], photo-responsiveness of the circadian clock[8] and emergence rhythms[9]. While these studies offer insights into how evolution can tinker with known circadian clock molecules, genome-wide association studies[10,11] and other forward genetic approaches (reviewed in[12]) are essential to provide a comprehensive, unbiased assessment of natural timing variation, for instance underlying human sleep-phase disorders. While the alleles underlying human sleep disorders embody disease states and the adaptive nature of human chronotypes is obscure, the chronotypes of *C. marinus* clearly represent evolutionary adaptations to their habitat. Our study aims to illustrate by which genes and mechanisms an organism can successfully adapt to its specific ecological "timing niche". In addition, the genetic dissection of adaptive natural variants of non-circadian rhythms[13], as also present in *C. marinus,* may provide an entry point into their unknown molecular mechanisms.

As a starting point for these analyses, we sequenced, assembled, mapped and annotated a *C. marinus* reference genome.

The *Clunio* genome and QTLs for timing

Our reference genome CLUMA_1.0 from the *Jean* laboratory strain contains 85.6 Mb of sequence (Table I), close to the previous flow-cytometry estimate of 95 Mb[6], underlining that chironomids have generally small genomes[14-16]. The final assembly has a scaffold N50 of 1.9 Mb. Genome-wide genotyping of a mapping family with Restriction-site Associated DNA (RAD) sequencing allowed anchoring of 92% of the reference sequence consistently along a genetic linkage map (Fig. 1a, Extended Data Fig. 2), improving the original linkage map (Supplementary Method 5). Automated genome annotation resulted in 21,672 gene models. Protein similarity and available transcripts support 14,041 gene models (Table S1), within the range of gene counts for *Drosophila melanogaster* (15,507) and *Anopheles gambiae* (13,460). Thus, the very small *C. marinus* genome appears to be complete (Table I; Extended Data Figure 2a; Supplementary Note 1; Table S2). The *C. marinus* reference genome makes chironomids the third dipteran subfamily with an annotated genome reconstructed to chromosome-scale (Fig. 1a, Extended Data Fig.2, 3b-f) and thus represents a valuable resource for comparative genomics.

To test the quality of our reference genome, we performed a basic genome characterization and comparison to other dipterans. We delineated the five *C. marinus* chromosome arms (Supplementary Note 2; Extended Data Fig. 3c; Table S3), homologized them to *D. melanogaster* and *A. gambiae* by synteny comparisons (Extended Data Fig. 3 and 4, Supplementary Note 2; Table S3),

found the ZW-like sex-linked locus in *C. marinus*[6] outside the X chromosome homolog (Supplementary Note 2), and detected an elevated rate of chromosomal re-arrangements (Fig. 1a; Supplementary Note 3; Extended Data Fig. 2, 3b-f, 4). Taken together, the *C. marinus* reference genome appears well assembled.

As the next step towards identifying the molecular basis of circadian and circalunar timing adaptations in *C. marinus*, we refined the previously identified timing QTL positions[6] based on the new high-density RAD markers (Table S4; Supplementary Note 4) and determined the reference sequence corresponding to the QTL confidence intervals (Fig. 1, orange and cyan bars; Table S4). None of the core circadian clock genes locates within the QTLs (Fig. 1a). Only *timeout/timeless2*, a *timeless* homolog with a minor role in circadian clock resetting[17], is located within the QTLs.

Genetic variation in *Clunio* timing strains

We then re-sequenced the *Por* and *Jean* strains (Extended Data Fig. 1), for which the initial QTL analysis was performed[6]. Two pools of 300 field-caught individuals were sequenced at >240x coverage (Table S5). Mapping reads against the reference genome identified 1,010,052 single nucleotide polymorphisms (SNPs), 72% of them being present in both the *Por* and *Jean* strains. Based on all SNPs we determined genetic differentiation (F_{ST}), genetic diversity (θ), and short-range linkage disequilibrium (LD; measured as r^2) (Fig. 1b; Extended Data Fig. 3c and 5a,b).

Genome-wide genetic differentiation between the *Por* and *Jean* strains is moderate (F_{ST} = 0.11), providing a good basis for screening the genome for local timing adaptation based on genetic divergence. According to QTL analysis, the

two circadian QTLs explain 85% of the daily timing difference, and the two circalunar QTLs explain the entire monthly timing difference (Table S4 and [6]). As each locus therefore has a strong effect on timing, selection against maladapted alleles must be strong and timing loci should be strongly differentiated.

Within the QTLs' confidence intervals, 158 SNPs and 106 indels are strongly differentiated (F_{ST}≥0.8; Fig. 1b; Extended Data Fig. 5; SNPs: red dots in F_{ST} panels, for genome-wide comparison see Supplementary Note 5,). We compiled a list of candidate genes for circadian and circalunar timing adaptations based on their proximity to differentiated SNPs and indels in the QTLs (Table S6). The candidate genes do neither comprise core circadian clock genes (*timeless2/timeout*: max. F_{ST} ≤ 0.5; average F_{ST} = 0.07), nor are they enriched for any particular pathway (GO-term analysis; Table S7).

Timing phenotype with genotype correlation

Given that the alleles responsible for timing adaptation likely originated from standing genetic variation (Supplementary Note 5), genetic variation at timing loci should not vary freely between strains, but rather strains with similar timing should share functionally relevant alleles. To identify such loci, we extended the genomic screen to three additional strains: *Vigo*, *Helgoland* (*He*) and *Bergen* (*Ber*; see Extended Data Fig. 1; Table S5 and S8). We then tested all five sequenced strains for correlations between genetic differentiation (F_{ST}) and timing differences, or geographic distances as a null model (Table S8).

Overall, genome-wide genetic differentiation is not correlated with circadian (r = 0.10, p = 0.31) or circalunar (r = 0.56, p = 0.12) timing differences, but with geographic distance ("isolation by distance"; r = 0.88, p = 0.008). Against this

genomic background signal of isolation by distance, we screened the genome in 5kb sliding windows for peaks of correlation between genetic differentiation and timing, resulting in a *correlation score* (Fig. 1b and Extended Data Fig. 5a,b, *CS* panels; 0 to 5; for details see Methods). Combining the evidence from the *Por* vs. *Jean* strain F_{ST} screen (Table S6) with these patterns of correlation between timing and genetic divergence reduced the candidate gene list to 49 genes (Table S9).

Particularly noteworthy, a single region in circadian QTL C2 is strikingly differentiated (Fig. 1b). In this region, LD in the *Por* strain is significantly elevated (permutation test; p = 0.002), and diversity significantly decreased in some stretches (permutation test; p = 0.037 and 0.020), compared to the *Por* genome average. This may indicate a recent episode of selection in *Por*, potentially during timing adaptation, as this region is also strongly enriched for timing-correlated polymorphisms (Fig. 1b, CS panel). The most extreme values of genetic differentiation, genetic diversity and timing correlation localize to the *Ca2+/Calmodulin dependent Kinase II.1* (*CaMKII.1*) locus and the anterior section of a gene homologous to the *big bang (bbg)* gene.

CaMKII affects the circadian core clock

The *CaMKII.1* locus not only harbors the highest number of differentiated polymorphisms (Table S9), but CaMKII has been shown to affect circadian timing. Mouse CaMKIIα phosphorylates CLOCK and facilitates its dimerization with BMAL *in vivo*[18]. An inactive CaMKIIα enzyme ("kinase-dead"-mutation; K42R) leads to dampened circadian rhythms, and a lengthened circadian free-running period[18]. CaMKII in *Drosophila* S2 cells also phosphorylates the CLOCK

protein[19], and inhibition of *Dme*-CaMKII in a sensitized background with reduced $[Ca^{2+}]$ levels lengthens the circadian free-running period[20], suggesting that the role of CAMKII in circadian timing is conserved across animals.

To verify if CaMKII can also affect the circadian core clock in *C. marinus*, we tested the effect of *Cma*-CaMKII.1 in a S2 cell-based assay[19,21]. We repeated previous experiments[19] showing that the chemical inhibition of endogenous *Drosophila* CaMKII reduces the amount of generated luciferase (Extended Data Fig. 6a), while addition of a $[Ca^{2+}]$-independent variant of CaMKII (mouse T286D) increases luciferase amounts (Extended Data Fig. 6b). Then we generated constructs for *C. marinus clock*, *C. marinus cycle*, as well as mutated kinase-dead (K42R) and $[Ca^{2+}]$-indepenent (T286D) versions of *Cma-CaMKII.1*. Transfection of *Cma-clock* and *Cma-cycle* into S2 cells leads to luciferase activity driven from the 3x69 per-promoter (Fig. 2a). The addition of $[Ca^{2+}]$-independent *Cma-CaMKII.1* leads to a significant increase in the luciferase signal (Fig. 2a), whereas addition of the kinase-dead *Cma-CaMKII.1* does not enhance luciferase activity (Fig. 2a). This set of experiments strongly suggests that CaMKII kinase activity enhances E-box dependent transcription via the CLOCK/CYCLE dimer in *C. marinus*.

CaMKII.1 splicing correlates with timing

But how can the polymorphisms in the *Cma-CaMKII.1* locus affect the enzyme? We found two *CaMKII.1* alleles: one in the early emerging *Por*, *He* and *Ber* strains, and another in the late emerging *Jean* and *Vigo* strains. Most strain-specific polymorphisms are located in introns (Fig. 2b,c; TableS9). If they are meaningful, they should affect *CaMKII.1* expression and/or splicing. *Cma-CaMKII.1* has four functional domains (Fig. 2b)[22]. The majority of differentiated polymorphisms

cluster in the region of the variable linker domain (compare Fig. 2b,c), including a 125bp insertion (red dot in Fig. 2c; Extended Data Fig. 7). We identified four alternatively spliced full-length transcripts of *C. marinus CaMKII.1* (RA-RD), which differ in the linker length (Fig. 2b). High-coverage RNA sequencing gave evidence for differential exon usage between the *Jean* and *Por* strains, as well as for previously non-annotated exons within the variable linker region (Extended Data Fig. 6c). PCR and Sanger sequencing confirmed several partial transcripts of additional splice variants of the linker region (RE to RO; Fig. 2b). We used transcript-specific qPCR to quantify all transcripts. Generally, transcripts RE to RO are very lowly expressed. Of those, only RO showed quantifiable expression differences between the *Jean* vs. *Por* strains (Fig. 3a, Extended Data Fig. 6d). Importantly, transcript-specific qPCR confirmed significant differential expression of the major transcripts in the *Jean* vs. *Por* strains (Fig. 3a, Extended Data Fig. 6d), matching the RNAseq data (Extended Data Fig. 6c). Consistently, variants with long linkers (RA, RB) are higher expressed in the *Por* strain and shorter variants (RD, RO) are higher expressed in the *Jean* strain (Fig. 3a, Extended Data Fig. 6c,d).

If the detected differences in *CaMKII.1* splice variant abundance are associated with the timing differences, they should be directly caused by the strain-specific polymorphisms at the *CaMKII.1* locus. In order to test this, we generated minigenes that contained the alternatively spliced linker region of the *CaMKII.1* locus from either the *Jean* or the *Por* strain. The two minigenes were transfected into *Drosophila* S2R+ cells and expression of splice variants was analyzed by radioactive RT-PCR (Fig. 3b,c). We detected four variants, corresponding to splice variants RB, RC, RD and RO. All variants show the same strain-specific

abundance differences in the S2R+ cell assay and in *C. marinus in vivo* (Fig. 3a,b). Since the cellular context is the same for both the *Jean* and *Por* minigenes in the S2R+ assay, *trans*-acting elements can be excluded as the cause of differential splicing, implying that it is a direct result of the genomic sequence differences at the *Cma-CaMKII.1* locus. While splice variants RB, RC and RD and their constituting exons are conserved in *D. melanogaster* (see Flybase annotations and [23]), a *D. melanogaster* RA counterpart does not exist. This may explain why this variant is undetectable in S2R+ cells.

From splice variants to timing differences

CaMKII linker-length variants have been investigated in several species. *D. melanogaster* CaMKII isoforms corresponding to the RB, RC and RD variants of *C. marinus*, have different substrate affinities and rates of target phosphorylation[23]. These activity differences are explained by the fact that CaMKII functions as a dodecamer, and the linker length determines the compactness and thus the substrate accessibility of the holoenzyme – enzymes with long linkers have higher activity. This structure-functional relationship is likely universal, as it is conserved between humans and *C. elegans*[22,24].

Inactivation or inhibition of CaMKII lengthens circadian period in mouse and fruit flies[18,20]. A connection between circadian period length and phase of activity in light/dark cycles is known from *per* mutations in *D. melanogaster*[25] and human chronotypes[26]. These findings imply that in *C. marinus* the more active and more readily [Ca^{2+}]-activated long-linker *CaMKII.1* variants should advance adult emergence by shortening the circadian clock period. Indeed, we find that the early emerging *Por* and *He* strains, which possess the same long-linker biased

CaMKII.1 alleles, have shorter free-running circadian clock periods than the late emerging *Jean* strain (Fig. 3d).

Integrating our results with those from the aforementioned literature, the scenario emerges that regulating the ratio of *CaMKII.1* splice variants constitutes an evolutionary mechanism to adapt circadian timing (Extended Data Fig.8): *CaMKII.1* genomic sequence differences lead to differential *CaMKII.1* splicing and activity. Among a number of possible targets this impacts on CLOCK/CYCLE dimer-dependent transcription, which in turn affects circadian period length and ultimately results in adult emergence time differences.

Discussion

Annual, lunar, and tidal rhythms, as well as natural timing variation between individuals, are important and widespread, yet poorly understood, phenomena. The comprehensively mapped *C. marinus* reference genome and the genetic variation panel for five strains with differing circadian and circalunar timing establish new resources to gain insight into these topics.

We identified *C. marinus* orthologs for all core circadian clock genes, none of which appears to be involved in circadian or circalunar timing adaptations. For circalunar timing, this supports the molecular independence of the circalunar clock from the circadian clock as reported for *Platynereis dumerilii*[27].

For circadian timing, strain-specific modulation in alternative splicing of *CaMKII.1* emerges as a likely mechanism for natural adaptation. In the light of previous experiments in *Drosophila* and mouse[18-20,23], it seems most likely that differences in CaMKII activity of the different splice forms lead to circadian timing differences via phosphorylation of CLOCK/CYCLE (Extended Data Fig. 8).

It is also conceivable that CaMKII affects circadian timing via other targets. For example, CaMKII is known to phosphorylate the cAMP response element binding protein (CREB)[28,29]. CREB is linked to the circadian clock by cAMP response elements (CRE) in the promoters of the *period* and *timeless* genes[30,31], and by physical interaction of the CREB binding protein (CBP) with CREB, CLOCK and CYCLE[32,33]. Furthermore, one of CaMKII's best-studied roles is the morphological modulation of neuronal plasticity and connectivity[34-36]. Such changes in connectivity have been increasingly implicated as part of the circadian timing mechanism in *Drosophila and* mammals[37]. Interestingly,

CaMKII's role in shaping neuronal connectivity has also been suggested to link to several neuropsychiatric diseases[38], which often co-occur with chronobiological disruptions[39-42]. Could the modulation of CaMKII activity constitute a molecular link between these phenomena?

References

1 Neumann, D. Die lunare und tägliche Schlüpfperiodik der Mücke *Clunio* - Steuerung und Abstimmung auf die Gezeitenperiodik. *Zeitschrift für Vergleichende Physiologie* **53**, 1-61 (1966).

2 Neumann, D. Temperature compensation of circasemilunar timing in the intertidal insect *Clunio*. *Journal of Comparative Physiology A - Sensory Neural and Behavioral Physiology* **163**, 671-676 (1988).

3 UKHO. *ADMIRALTY Tide Tables*. (2014).

4 Neumann, D. Genetic adaptation in emergence time of *Clunio* populations to different tidal conditions. *Helgoländer wissenschaftliche Meeresuntersuchungen* **15**, 163-171 (1967).

5 Kaiser, T. S., Neumann, D. & Heckel, D. G. Timing the tides: Genetic control of diurnal and lunar emergence times is correlated in the marine midge *Clunio marinus*. *BMC Genetics* **12**, 49, doi:10.1186/1471-2156-12-49 (2011).

6 Kaiser, T. S. & Heckel, D. G. Genetic Architecture of Local Adaptation in Lunar and Diurnal Emergence Times of the Marine Midge *Clunio marinus* (Chironomidae, Diptera). *PLoS ONE* **7**, e32092, doi:10.1371/journal.pone.0032092 (2012).

7 Sawyer, L. A. *et al.* Natural Variation in a *Drosophila* Clock Gene and Temperature Compensation. *Science* **278**, 2117-2120, doi:10.1126/science.278.5346.2117 (1997).

8 Sandrelli, F. *et al.* A Molecular Basis for Natural Selection at the *timeless* Locus in *Drosophila melanogaster*. *Science* **316**, 1898-1900 (2007).

9 Pegoraro, M. *et al.* Molecular Evolution of a Pervasive Natural Amino-Acid Substitution in *Drosophila cryptochrome*. *PLoS ONE* **9**, e86483, doi:10.1371/journal.pone.0086483 (2014).

10 Lane, J. M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat Commun* **7**, doi:10.1038/ncomms10889 (2016).

11 Hu, Y. *et al.* GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat Commun* **7**, doi:10.1038/ncomms10448 (2016).

12 Jones, C. R., Huang, A. L., Ptáček, L. J. & Fu, Y.-H. Genetic basis of human circadian rhythm disorders. *Experimental Neurology* **243**, 28-33, doi:http://dx.doi.org/10.1016/j.expneurol.2012.07.012 (2013).

13 Tessmar-Raible, K., Raible, F. & Arboleda, E. Another place, another timer: Marine species and the rhythms of life. *Bioessays* **33**, 165-172, doi:10.1002/bies.201000096 (2011).

14 Gusev, O. *et al.* Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nature Communications* **5**, doi:10.1038/ncomms5784 (2014).

15 Cornette, R. *et al.* Chironomid midges (Diptera, Chironomidae) show extremely small genome sizes. *Zoological Science* **32**, 248-254, doi:10.2108/zs140166 (2015).

16 Kelley, J. L. *et al.* Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nature Communications* **5**, doi:10.1038/ncomms5611 (2014).

17 Benna, C. *et al.* Drosophila timeless2 Is Required for Chromosome Stability and Circadian Photoreception. *Current Biology* **20**, 346-352, doi:10.1016/j.cub.2009.12.048 (2010).

18 Kon, N. *et al.* CaMKII is essential for the cellular clock and coupling between morning and evening behavioral rhythms. *Genes & Development* **28**, 1101-1110, doi:10.1101/gad.237511.114 (2014).

19 Weber, F., Hung, H. C., Maurer, C. & Kay, S. A. Second messenger and Ras/MAPK signalling pathways regulate CLOCK/CYCLE-dependent transcription. *Journal of Neurochemistry* **98**, 248-257, doi:10.1111/j.1471-4159.2006.03865.x (2006).

20 Harrisingh, M. C., Wu, Y., Lnenicka, G. A. & Nitabach, M. N. Intracellular Ca2+ regulates free-running circadian clock oscillation in vivo. *Journal of Neuroscience* **27**, 12489-12499, doi:10.1523/jneurosci.3680-07.2007 (2007).

21 Nawathean, P. & Rosbash, M. The doubletime and CKII kinases collaborate to potentiate Drosophila PER transcriptional repressor activity. *Molecular Cell* **13**, 213-223, doi:Doi 10.1016/S1097-2765(03)00503-3 (2004).

22 Chao, L. H. *et al.* A Mechanism for Tunable Autoinhibition in the Structure of a Human Ca2+/Calmodulin-Dependent Kinase II Holoenzyme. *Cell* **146**, 732-745, doi:10.1016/j.cell.2011.07.038 (2011).

23 GuptaRoy, B. *et al.* Alternative splicing of *Drosophila* calcium/calmodulin-dependent protein kinase II regulates substrate specificity and activation. *Molecular Brain Research* **80**, 26-34, doi:Doi 10.1016/S0169-328x(00)00115-7 (2000).

24 Chao, L. H. *et al.* Intersubunit capture of regulatory segments is a component of cooperative CaMKII activation. *Nat Struct Mol Biol* **17**, 264-272, doi:http://www.nature.com/nsmb/journal/v17/n3/suppinfo/nsmb.1751_S1.html (2010).

25 Hamblen-Coyle, M. J., Wheeler, D. A., Rutila, J. E., Rosbash, M. & Hall, J. C. Behavior of period-altered circadian rhythm mutants of Drosophila in light: Dark cycles (Diptera: Drosophilidae). *J Insect Behav* **5**, 417-446, doi:10.1007/bf01058189 (1992).

26 Brown, S. A. *et al.* Molecular insights into human daily behavior. *Proceedings of the National Academy of Sciences* **105**, 1602-1607, doi:10.1073/pnas.0707772105 (2008).

27 Zantke, J. *et al.* Circadian and Circalunar Clock Interactions in a Marine Annelid. *Cell Reports* **5**, 99-113 (2013).

28 Sun, P. Q., Enslen, H., Myung, P. S. & Maurer, R. A. Differential Activation of Creb by Ca2+/Calmodulin-Dependent Protein-Kinases Type-II and Type-IV Involves Phosphorylation of a Site That Negatively Regulates Activity. *Genes & Development* **8**, 2527-2539, doi:Doi 10.1101/Gad.8.21.2527 (1994).

29 Wu, X. L. & McMurray, C. T. Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the CREB-binding protein. *Journal of Biological Chemistry* **276**, 1735-1741, doi:Doi 10.1074/Jbc.M006727200 (2001).

30 Belvin, M. P., Zhou, H. & Yin, J. C. P. The Drosophila dCREB2 gene affects the circadian clock. *Neuron* **22**, 777-787, doi:Doi 10.1016/S0896-6273(00)80736-9 (1999).

31 Okada, T. *et al.* Promoter analysis for daily expression of Drosophila timeless gene. *Biochem Bioph Res Co* **283**, 577-582, doi:Doi 10.1006/Bbrc.2001.4793 (2001).

32 Lim, C. *et al.* Functional role of CREB-binding protein in the circadian clock system of Drosophila melanogaster. *Molecular and Cellular Biology* **27**, 4876-4890, doi:10.1128/MCB.02155-06 (2007).

33 Lee, Y. *et al.* Coactivation of the CLOCK-BMAL1 complex by CBP mediates resetting of the circadian clock. *J Cell Sci* **123**, 3547-3557, doi:10.1242/jcs.070300 (2010).

34 Kalil, K., Li, L. & Hutchins, B. I. Signaling mechanisms in cortical axon growth, guidance and branching. *Frontiers in Neuroanatomy* **5**, doi:10.3389/fnana.2011.00062 (2011).

35 Hell, J. W. CaMKII: Claiming Center Stage in Postsynaptic Function and Organization. *Neuron* **81**, 249-265, doi:10.1016/j.neuron.2013.12.024 (2014).

36 McVicker, D. P., Millette, M. M. & Dent, E. W. Signaling to the microtubule cytoskeleton: An unconventional role for CaMKII. *Developmental Neurobiology* **75**, 423-434, doi:10.1002/dneu.22227 (2015).

37 Bosler, O., Girardet, C., Franc, J.-L., Becquet, D. & François-Bellan, A.-M. Structural plasticity of the circadian timing system. An overview from flies to mammals. *Frontiers in Neuroendocrinology* **38**, 50-64, doi:http://dx.doi.org/10.1016/j.yfrne.2015.02.001 (2015).

38 Robison, A. J. Emerging role of CaMKII in neuropsychiatric disease. *Trends in Neurosciences* **37**, 653-662, doi:http://dx.doi.org/10.1016/j.tins.2014.07.001 (2014).

39 Wulff, K., Gatti, S., Wettstein, J. G. & Foster, R. G. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat Rev Neurosci* **11**, 589-599, doi:http://dx.doi.org/10.1038/nrn2868 (2010).

40 Levandovski, R. *et al.* Depression Scores Associate With Chronotype and Social Jetlag in a Rural Population. *Chronobiology International* **28**, 771-778, doi:10.3109/07420528.2011.602445 (2011).

41 Zordan, M. A. & Sandrelli, F. Circadian clock dysfunction and psychiatric disease: could fruit flies have a say? *Frontiers in Neurology* **6**, doi:10.3389/fneur.2015.00080 (2015).

42 Logan, R. W. *et al.* Chronic Stress Induces Brain Region-Specific Alterations of Molecular Rhythms that Correlate with Depression-like Behavior in Mice. *Biological Psychiatry* **78**, 249-258, doi:http://dx.doi.org/10.1016/j.biopsych.2015.01.011 (2015).

43 Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. The Monarch Butterfly Genome Yields Insights into Long-Distance Migration. *Cell* **147**, 1171-1185, doi:10.1016/j.cell.2011.09.052 (2011).

44 Richards, S. *et al.* The genome of the model beetle and pest Tribolium castaneum. *Nature* **452**, 949-955, doi:10.1038/nature06784 (2008).

45 Weinstock, G. M. *et al.* Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443**, 931-949, doi:10.1038/nature05260 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments

We thank the members of the Tessmar-Raible, Raible, von Haeseler groups for discussions, S. Bannister and F. Raible for comments on the manuscript. The research leading to these results has received funding from the research platform “Rhythms of Life” of the University of Vienna to KT-R, TH and AvH, the FWF (http://www.fwf.ac.at/) START award (#AY0041321), the HFSP (http://www.hfsp.org/) research grant (#RGY0082/2010), the ERC (FP7/2007-2013)/ERC Grant Agreement 337011 to KT-R and the DFG grant (#HE5398/4-1) to FH. TSK was supported by the Vienna International PostDoctoral Program for Molecular Life Sciences (VIPS), L-TN and AvH by the University of Vienna Initiativkolleg I059-N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the study, interpreted the data: TSK, KT-R, AvH; field sampling, chronobiological experiments, (super)scaffolding, genetic/QTL mapping, population genomics: TSK; assembly to contigs: L-TN, TN, TSK; assembly filtering: TSK, TN; gap closing and repeated edge removal: TSK, FJS; RNAseq: TSK, FJS, HV; cDNA library: HV; genome annotation: TSK, DS; analysis of genome completeness: TSK, TN; chromosome homology and synteny comparisons: TSK, AZ; estimation of linkage disequilibrium: AJB; SNP effects and GO term analysis: DS, TSK; *Cma-CaMKII.1* analyses: BP, TSK, SD; minigene assay: MP, FH; contributed material: TH; wrote the manuscript: TSK, KT-R.

Author Information

All sequence data are deposited in the European Nucleotide Archive (ENA) under PRJEB8339. The reference genome is also on *ClunioBase* (http://cluniobase.cibiv.univie.ac.at). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials: K.T-R. (kristin.tessmar@mfpl.ac.at) and T.S.K. (kaiser@evolbio.mpg.de).

Fig. 1 Identification of candidate regions in the timing QTLs by combined genetic and molecular maps.

(a) The three linkage groups of *C. marinus* with reference scaffolds (right) anchored on a genetic linkage map (left). Scaffolds ordered and oriented: black bars; not oriented: grey bars; neither ordered nor oriented: white bars. Grey shadings: large non-recombining regions. QTLs: circadian (orange), circalunar (cyan). One circadian and circalunar QTL overlap, resulting in three physical QTL regions. **(b)** Population genomic analysis of QTL-C2. Panels 1-3: *Por* vs. *Jean* strains (blue vs. red in panel 2,3). *Panel 1*: Genetic differentiation, *Panel 2:* Genetic diversity (θ) in 20-kb (thin line) and 200-kb (thick line) windows. *Panel 3:* Linkage disequilibrium (r^2). *Panel 4:* Correlation Score (CS; 0 to 5) for genetic differentiation with circadian timing (top), circalunar timing (middle) and geographic distance (bottom) for *Vigo*, *Jean*, *Por*, *He*, *Ber* strains. Bottom numbers: scaffold IDs. For further details, including QTLs C1/L1 and L2, see Extended Data Fig. 5a,b.

Fig. 2 CaMKII.1 regulates CLK/CYC transcriptional activity and exhibits strain specific splice variants

(a) Additional *C. marinus* CaMKII.1 increases the transcriptional activity of *C. marinus* Clk and Cyc in a S2 cell luciferase assay using the 3x69 E-box containing enhancer[21]. (means; error bars: S.E.M.; two-sided Welch Two Sample t-test; biological replicates (BR): n=5, except for no *clk* control n=3; each BR represents the average of three prep replicates, *** $p<0.0005$). **(b)** Exons of full (RA-RD) and partial (RE-RO) *Cma-CaMKII.1* transcripts. **(c)** Distribution of SNPs (black), indels (orange) and a 125bp-insertion (red dot), all with $F_{ST} \geq 0.8$.

Fig. 3 *CaMKII.1* splicing depends on the genomic sequence variations and correlate with endogenous circadian period lengths

(a) qPCR values for *CaMKII.1* splice variants from *Por* vs. *Jean,* normalized to *Por* (non-normalized: Extended Data Fig. 6d); BR: *Por* n=9, *Jean* n=10; RO: *Por* n=3, *Jean* n=8, since RO was not detectable in six *Por* BRs, suggesting an even larger expression difference; means; error bars: S.E.M.; two-sided Wilcoxon rank sum test; *p<0.05; **p<0.005; ***p<0.0005, "ns" non-significant; Holm correction for multiple testing. (quantitative RNAsequencing of further BRs: Extended Data Fig. 6c) **(b)** Differential splicing of the *CaMKII.1* linker region in S2R+ cells, normalized to *Por*, BR: n=7; two-sided Two Sample T-test, otherwise as (a). **(c)** Representative phosphorimaging gel sections as quantified for (b), two separate lanes from the same gel (full gel: source data). **(d)** Free-running rhythm of adult emergence under constant dim white light (~100 lux). *He* and *Por* share *CaMKII.1* alleles, while *Jean* posses the other. Free-running period: time between subsequent emergence peaks was averaged, weighting each peak by the number of individuals.

Table I

	Clunio marinus	*Danaus plexippus*[43]	*Tribolium castaneum*[44]	*Apis mellifera*[45]
Total bases (Mbp)	86	278	160	236
Mapped sequence (%)	92	NA	90	79
Scaffold N50 (Mbp)	1.9	0.2	1	0.4
Contig N50 (kbp)	79	50	41	41
AT content (%)	68.3	68.4	67	66
Completeness (%)	98.0	98.5	NA	NA
Platform	Illumina	Illumina + 454	Sanger + BAC	Sanger + BAC

Comparison of the *C. marinus* reference assembly with published model insect genomes

Machine readable superscaffolding data and the computer source code for the removal of repeated edges are supplied as source data files.

Online Methods

Animal culture and light regimes

The *Clunio marinus* laboratory stocks were bred according to Neumann[1], care was provided by the MFPL aquatic facility. Briefly, they were kept in 20x20x5cm plastic containers with sand and natural seawater diluted to 15‰ with desalted water, fed diatoms (*Phaeodactylum tricornutum*, strain UTEX 646) in early larval stages and nettle powder in later stages. Temperature in the climate chambers was set to 20°C and the light dark cycle (LD) was 12:12 (except where noted differently). Moonlight was simulated with an incandescent flashlight bulb (about 1 Lux), which was switched on all night for four successive nights every 30 days.

Genome assembly

The genome assembly process (Extended Data Fig. 9a) was based on three sequencing libraries (Table S10): A 0.2kb insert library was prepared from a single adult male of the *Jean* laboratory strain (established from field samples taken at *St. Jean-de-Luz*, France, in 2007; >12 generations in the laboratory), which was starved and kept in seawater with Penicillin (60 units/ml), Streptomycin (60 µg/ml) and Neomycin (120 µg/ml) during the last 2 weeks of development. DNA was extracted with a salting out method[46], sheared on a *Covaris S2* sonicator (frequency sweeping mode; 4°C; duty cycle: 10%; intensity: 7; cycles/burst: 300; microTUBE AFA Fiber 6x16 mm; 30 s) and prepared for Illumina sequencing with standard protocols. A 2.2kb and a 7.6kb insert library were prepared from a polymorphic DNA pool of >300 field-caught *Jean* adult males by Eurofins MWG Operon (Ebersberg, Germany) according to their

551 proprietary protocol. Each library was sequenced in one lane of an Illumina
552 HiSeq2000 with 100bp paired-end reads at the Next Generation Sequencing unit
553 of the Vienna Biocenter Core Facilities (VBCF; http://vbcf.ac.at).
554 Reads were filtered for read quality, adapter and spacer sequences with
555 *cutadapt*[47] (–b 4 –n 3 –e 0.1 –O 8 –q 20 –m 13) and deduplicated with *fastq-mcf*
556 from *ea-utils*[48] (–D 70). Read pairs were interleaved with *ngm-utils*[49], leaving
557 only paired reads. Contamination with human DNA found in the 0.2kb library
558 was removed by deleting reads matching the human genome at a phred-scaled
559 quality score ≥20 (alignment with BWA[50]).
560 Assembly into contigs with *Velvet*[51] (scaffolding disabled; 57bp kmers as
561 determined by *VelvetOptimiser*[52]) was based solely on the less polymorphic
562 0.2kb library. About 600 remaining adaptor sequences at the ends of assembled
563 contigs were trimmed with *cutadapt* (-O 8 -e 0.1 -n 3). For assembly statistics see
564 Table S11.
565 Scaffolding of the contigs was based on all three libraries and performed with
566 SSPACE[53] in two iterations, i.e. scaffolds from the first round were scaffolded
567 again. Using different parameters in the iterations (Table S12) allowed different
568 connections to be made and thus increased scaffold connectivity (Table S13).
569 The effect is likely due to the polymorphic nature of the 2.2kb and 7.6kb
570 libraries; it results in a "population-consensus most common arrangement of the
571 scaffolds". The iterative scaffolding process was performed with and without
572 applying a size cutoff excluding contigs <1kb, resulting in two independent
573 assemblies (CLUMA_0.3 and CLUMA_0.4; see Extended Data Fig. 9a), which
574 differed in overall connectivity and sequence content (Table S11), but also in the
575 identity and structure of the large scaffolds. In order to combine both

connectivity and sequence content, and in order to resolve the contradictions in the structure of the largest scaffolds, the two assemblies were compared and reconciled in a manual super-scaffolding process, as detailed in Supplementary Method 1. Briefly, the overlap of scaffolds from the two assemblies was tested with BLAST searches and represented in a graphical network structure. Scaffolds with congruent sequence content in both assemblies would result in a linear network, whereas scaffolds with contradictory sequence content would result in branching networks. At the same time, both assemblies were subject to genetic linkage mapping based on genotypes obtained from Restriction-site Associated DNA sequencing (RAD sequencing) of a published mapping family[6] (Supplementary Method 2). The resulting genetic linkage information served to resolve the branching networks into the longest possible unambiguous linear sub-networks with consistent genetic linkage information (see scheme A in Supplementary Method 1). Finally, the structure of the resulting super-scaffolds was coded in YAML format and translated into DNA sequence with *Scaffolder*[54], resulting in 75 mapped super-scaffolds.

The remaining small and unmapped scaffolds were filtered for fragments of the mitochondrial genome, the histone gene cluster and 18S/28S ribosomal rDNA gene cluster, which were assembled separately (Supplementary Method 3; Extended Data Fig. 10). Unmapped scaffolds were also filtered for obvious contamination from other species (Supplementary Method 3). The degree to which the remaining unmapped scaffolds are leftover polymorphic variants of parts of the mapped super-scaffolds was estimated by blasting the former against the latter (Supplementary Method 3; Table S14).

All scaffolds were subject to gap closing with *GapFiller*[55] and repeated edges, i.e. gaps with repetitive sequence at both sides that are generally due to genetic polymorphism, were assessed and if possible removed with a custom script (Supplementary Method 4; code available supplied as Source Data File).

The final assembly CLUMA_1.0 was submitted under project PRJEB8339 (75 mapped scaffolds; 23,687 unmapped scaffolds >=100bp). The assembly and further information can also be obtained from *ClunioBase* (http://cluniobase.cibiv.univie.ac.at)

Reconstruction of chromosomes and QTL analysis

Genetic linkage information for the final 75 super-scaffolds was obtained by repeating read mapping to genotype calling for the RAD sequencing experiment as described above (Supplementary Method 2), but now with assembly CLUMA_1.0 as a reference. This allowed to place and orient super-scaffolds along the genetic linkage map (Fig. 1a, Extended Data Fig.2). The positions of the recombination events within a scaffold were approximated as the middle between the positions of the two RAD markers between which the marker pattern changed from one map location to the next. The published genetic linkage map was refined and revised (Supplementary Method 5; Extended Data Fig. 2). Based on the refined linkage map, QTL analysis of the published mapping family was repeated as described[6] (Table S4; Supplementary Note 5). Using the correspondence between the reference assembly and the genetic linkage map, we were able to directly identify the genomic regions corresponding to the QTLs' confidence intervals (Fig. 1, Extended Data Fig. 5a,b).

Transcript sequencing

From previous experiments assembled transcripts were available from a normalized cDNA library of all life stages and various *C. marinus* strains (454 sequencing) and RNA sequencing data was available for *Jean* strain adults (Illumina sequencing). Furthermore, specifically for genome annotation, RNA from 80 third instar larvae each from the *Jean* and *Por* laboratory strains was prepared for RNA sequencing according to standard protocols (Supplementary Method 6). Each sample was sequenced on a single lane of an Illumina HiSeq 2000. All transcript reads were submitted to the European Nucleotide Archive (ENA) under project PRJEB8339.

For the adult and larval RNA sequencing data, raw reads were quality checked with *fastqc*[56], trimmed for adaptors quality with *cutadapt*[47] and filtered to contain only read pairs using the interleave command in *ngm-utils*[49]. Reads were assembled separately for larvae and adults with *Trinity*[57] (path_reinforcement_distance: 25; maximum paired-end insert size: 1500 bp; otherwise default parameters).

Genome annotation

Automated annotation was performed with MAKER2[58]. Repeats were masked based on all available databases in *repeatmasker*. MAKER2 combined evidence from assembled transcripts (see above), mapped protein datasets from *Culex quinquefasciatus* (CpipJ1), *Anopheles gambiae* (AgamP3), *Drosophila melanogaster* (BDGP5), *Danaus plexippus* (DanPle_1.0), *Apis mellifera* (Amel4.0), *Tribolium castaneum* (Tcas3), *Strigamia maritima* (Smar1) and *Daphnia pulex* (Dappu1) and *ab initio* gene predictions with AUGUSTUS[59] and SNAP[60] into gene models. AUGUSTUS was trained for *C. marinus* based on assembled transcripts from the normalized cDNA library. SNAP was run with parameters for *A.*

649 *mellifera*, which had the highest congruence with known *C. marinus* genes in
650 preliminary trials (Supplementary Method 7). MAKER was set to infer gene
651 models from all evidence combined (not transcripts only) and gene predictions
652 without transcript evidence were allowed. Splice variant detection was enabled,
653 single-exon genes had to be larger than 250bp and intron size was limited to a
654 maximum of 10 kb.
655 All gene models within the QTL confidence intervals, as well as all putative
656 circadian clock genes and light receptor genes were manually curated: Exon-
657 intron boundaries were corrected according to transcript evidence (~500 gene
658 models), chimeric gene models were separated into the underlying individual
659 genes (~100 gene models separated into ~300 gene models) and erroneously
660 split gene models were joined (~15 gene models). Finally, this resulted in 21,672
661 gene models, which were given IDs from CLUMA_CG000001 to
662 CLUMA_CG021672 ("CLUMA" for *Clunio marinus*, following the controlled
663 vocabulary of species from the UniProt Knowledgebase; CG for "computated
664 gene"). Splice variants of the same gene (detected in 752 gene models) were
665 identified by the suffix "-RA", "-RB", etc. and the corresponding proteins by the
666 suffix "-PA", "-PB", etc..
667 Gene models were considered as supported if they overlapped with mapped
668 transcripts or protein data (Table S1). Gene counts for *Drosophila melanogaster*
669 were retrieved from BDGP 5, version 75.546 and for *Anopheles gambiae* from
670 AgamP3, version 75.3. The putative identities of the *C. marinus* gene models
671 were determined in reciprocal BLAST searches, first against UniProtKB/Swiss-
672 Prot (8,379 gene models assigned) and if no hit was found against nr at NCBI
673 (1,802 additional genes assigned). Reciprocal best hits at an e-value < 1*e-10

were considered putative orthologs (termed "putative gene X"), non-reciprocal hits at the same e-value were considered paralogs (termed "similar to"). All remaining gene models were searched against the PFAM database of protein domains (111 gene models assigned; termed "gene containing domain X"). If still no hit was found, the gene models were left unassigned ("NA").

Synteny comparisons

Genome-wide synteny between the *C. marinus*, *D. melanogaster* and *A. gambiae* genomes was assessed based on reciprocal best BLAST hits (e-value < $10*e^{-10}$) between the three protein datasets (Ensembl Genomes, Release 22, for *D. melanogaster* and *A. gambiae*). Positions of pairwise orthologous genes were retrieved from the reference genomes (BDGP 5, AgamP3 and CLUMA_1.0) and plotted with Circos[61]. *C. marinus* chromosome arms were delimited based on centromeric and telomeric signatures in genetic diversity and linkage disequilibrium (Extended Data Fig. 3c; Table S3; for data source see "strain re-sequencing" below). Homologies for *C. marinus* chromosome arms were assigned based on enrichment with putative orthologous genes from specific chromosome arms in *D. melanogaster* and *A. gambiae* (Extended Data Figures 3,4; Table S3). Additionally, for the 5,388 detected putative 1:1:1 orthologs, microsynteny was assessed by testing if all pairs of directly adjacent genes in one species were also directly adjacent in the other species. The degree of microsynteny was then calculated as the fraction of conserved adjacencies among all pairs of adjacent genes. From this fraction the relative levels of chromosomal rearrangements in the evolutionary lineage leading to *C. marinus* were estimated (Supplementary Note 2; Extended Data Fig. 4).

Strain re-sequencing

699 Genetic variation in five *C. marinus* strains (Extended Data Fig. 1) was assessed
700 based on pooled-sequencing data from field-caught males from the strains of St.
701 Jean-de-Luz (*Jean*, Basque Coast, France; sampled in 2007; n=300), Port-en-
702 Bessin (*Por*, Normandie, France; 2007; n=300), as well as Vigo (Spain; 2005;
703 n=100), Helgoland (*He*, Germany; 2005; n=300) and Bergen (*Ber*, Norway; 2005;
704 n=100). Samples from Vigo and Bergen, were provided by Dietrich Neumann and
705 Christina Augustin, respectively. For each strain we chose the largest available
706 number of individuals to get the best possible resolution of allele frequencies.
707 Females are not available, because they are virtually invisible in the field. For an
708 overview of the experimental procedure, see Extended Data Fig. 9b. DNA was
709 extracted with a salting out method[46] from sub-pools of 50 males, the DNA pools
710 were mixed at equal DNA amounts, sheared and prepared as described above
711 and sequenced on four lanes of an Illumina HiSeq2000 with paired-end 100 bp
712 reads (*Ber* and *Vigo* combined in one lane, distinguished by index reads). All
713 reads were submitted to the European Nucleotide Archive (ENA) under project
714 PRJEB8339. Sequencing reads were filtered for read quality and adapter
715 sequences with *cutadapt*[47] (–b –n 2 –e 0.1 –O 8 –q 13 –m 15), interleaved with
716 *ngm-utils*[49] and deduplicated with *fastq-mcf* from *ea-utils*[48] (–D 70). Reads were
717 aligned to the mapped super-scaffolds of assembly CLUMA_1.0 with *BWA*[50] (*aln*
718 and *sampe*; maximal insert size (bp): –a 1500).

719 **Detection of re-arrangements**

720 Based on the unfiltered alignments, the samples from *Por* and *Jean* were
721 screened for genomic inversions and insertion-deletions relative to the reference
722 sequence with the multi-sample version of DELLY[62]. Paired-end information was

723 only considered if the mapping quality was high (q≥20) (see also Supplementary
724 Note 4).

725 **Population genomic analysis of the timing strains**

726 For population genomic analysis (Extended Data Fig. 9b), the alignments of the
727 pool-seq data from *Vigo*, *Jean*, *Por*, *He* and *Ber* were filtered for mapping quality
728 (q≥20), sorted, merged and indexed with SAMtools[63]. Reads were re-aligned
729 around indels with the *RealignerTargetCreator* and the *IndelRealigner* in *GATK*[64].
730 The resulting coverage per strain is given in Table S5.

731 For identification of single nucleotide polymorphisms (SNPs), a pileup file was
732 created with the *mpileup* command of SAMtools[63]. Base Alignment Quality (BAQ)
733 computation was disabled (–B); instead, after creating a synchronized file with
734 the *mpileup2sync* script in PoPoolation2[65], indels that occurred more than ten
735 times were masked (including 3bp upstream and downstream) with
736 PoPoolations2's *identify-indel-regions* and *filter-sync-by-gtf* scripts. F_{ST} values
737 were determined with the *fst-sliding* script of *PoPoolation2*, applying a minimum
738 allele count of 10 (so that any false-positive SNPs resulting from the remaining
739 unmasked indels were effectively excluded) and a minimum coverage of 40x for
740 the *Por* vs. *Jean* comparison or 10x for the comparison of all five strains. F_{ST} was
741 calculated at single base resolution, as well as in windows of 5kb (step size: 1kb).
742 Individual SNPs were only considered for further analyses or plotted if they were
743 significantly differentiated as assessed by Fisher's exact test (*fisher-test* in
744 *PoPoolation2*).

745 Average genome-wide genetic differentiation between timing strains, as
746 obtained by averaging over 5kb sliding-windows, was compared to the
747 respective timing differences and geographic distances (see Table S8) in Mantel

748 tests (Pearson's product moment correlation; 9,999 permutations), as
749 implemented in the *vegan* package in the R statistical programming environment
750 R[66]. Geographic distances and circadian timing differences were determined as
751 described previously[67] (see Table S8). For determination of lunar timing
752 differences when comparing lunar with semilunar rhythms see Supplementary
753 Note 6. In order to find genomic regions for which genetic differentiation is
754 correlated with the timing differences between strains, the Mantel test was then
755 applied to 5kb genomic windows every 1kb along the reference sequence. 5kb is
756 roughly the average size of a gene locus in *C. marinus*. Windows with a
757 correlation coefficient of $r \geq 0.5$ were tested for significance (999 permutations).
758 For each genomic position the number of overlapping significantly correlated
759 5kb windows was enumerated, resulting in a correlation score (CS; ranging from
760 0 to 5).
761 Genetic diversity, measured as Watterson's theta (θ_W), for each strain was
762 assessed with *PoPoolation1.1.2*[68] in 20kb windows at 10kb steps. In order to save
763 computing time, the pileup files of *Jean*, *Por* and *He* were linearly downscaled to
764 100x coverage with the *subsample-pileup* script ("fraction" option), positions
765 below 100x coverage being discarded. Indel regions were excluded (default in
766 *PoPoolation 1.1.2*) and a minimum of 66% of a sliding window needed to be
767 covered. SNPs were only considered in θ_W calculations if present ≥2 times,
768 leading to slight inconsistencies in θ_W estimates between strains due to differing
769 coverage, but not affecting diversity comparisons within strains.
770 Linkage disequilibrium between the SNPs was determined for the *Por* and *Jean*
771 strains with LDx[69], assuming physical linkage between alleles on the same read
772 or read pairs. r^2 was determined by a maximum likelihood estimator, minimum

773 and maximum read depths corresponded to the 2.5% and 97.5% coverage
774 depths for each population (*Jean:* 111 to 315, *Por*: 98 to 319), total insert
775 distance was limited to 600bp, minimum phred-scaled base quality was 20,
776 minimum allele frequency was 0.1 and a minimum coverage per pair of SNPs
777 was 11. SNPs were binned by their physical distance for the plots (0-200bp, 200-
778 400bp, 400-600bp), with the mean value plotted.
779 Finally, small indels (<30bp) in the *Por* and *Jean* strains were detected with the
780 *UnifiedGenotyper* (–glm INDEL) in *GATK*[64] for positions with more than 20x
781 coverage. Genetic differentiation for indels was calculated with the classical
782 formula $F_{ST} = (H_T - H_S) / H_T$, where H_S is the average expected heterozygosity
783 according to Hardy-Weinberg Equilibrium (HWE) in the two subpopulations and
784 H_T is the expected heterozygosity in HWE of the hypothetical combined total
785 population. If more than two alleles were present, only the two most abundant
786 alleles were considered in the calculation of F_{ST}.
787 **Assessment of candidate genes**
788 Gene models from the automated annotation were considered candidate genes, if
789 they fulfilled the following criteria: (1) The gene was located within the
790 reference sequence corresponding to the QTL confidence intervals as
791 determined for the *Por* and *Jean* strains. (2) The gene contained a strongly
792 differentiated SNP or small indel or they were directly adjacent to such a SNP or
793 small indel ($F_{ST} \geq 0.8$ for *Por* vs. *Jean*, i.e. the strains used in QTL mapping). This
794 resulted in a preliminary list of 133 genes based on the *Por* vs. *Jean* comparison
795 (Table S6). These candidate genes were narrowed down based on their overlap
796 with genomic 5kb windows, for which genetic differentiation between five

797 European timing strains correlated with their timing differences (Fig. 1a;
798 Extended Data Fig. 5a,b; Table S9).
799 The location and putative effects of the SNPs and indels relative to the gene
800 models were assessed with SNPeff[70] (–ud 0, otherwise default parameters;
801 Extended Data Fig. 5c,d; Table S6 and S9).
802 For Gene Ontology (GO) term analysis, all *C. marinus* gene models with putative
803 orthologs in the UniProtKB/Swiss-Prot and nr databases based on reciprocal
804 best BLAST hits (see above) were annotated with the GO terms of their detected
805 orthologs (6.837 gene models). Paralogs were not annotated. The enrichment of
806 candidate SNPs and indels (F_{ST}≥0.8 between *Por* and *Jean*) in specific GO terms
807 was tested with SNP2GO[71] (min.regions=1, otherwise default parameters).
808 Hyper-geometric sampling was applied to test if individual genes of a GO term or
809 a whole pathway of genes are enriched for SNPs (Table S7).
810 **Molecular characterization of CaMKII.1**
811 RNAseq data of the *Por* and *Jean* strains for *CaMKII.1* were obtained from the
812 larval RNA sequencing experiment described above. Besides four assembled full-
813 length transcripts (RA to RD) from RNAseq and assembled EST libraries,
814 additional partial transcripts (RE to RO) were identified by PCR amplification
815 (for PCR primers see Table S15), gel extraction (QIAquick Gel Extraction Kit,
816 Qiagen), cloning with the CloneJET PCR Cloning Kit (Thermo Scientific) and
817 Sanger sequencing with pJET1.2 primers (LGC Genomics & Microsynth). cDNA
818 was prepared from RNA extracted from LIII larvae of the *Por* and *Jean* laboratory
819 strains (RNA extraction with RNeasy Plus Mini Kit, Qiagen; reverse transcription
820 with QuantiTect Reverse Transcription Kit, Qiagen).

qPCR was performed with variant-specific primers and actin as control gene (Table S16). cDNA was obtained from independent pools of 20 third instar larvae of the *Por* and *Jean* strains. Sample size was ten per strain to cover different time points during the day and to test for reproducibility (two samples each at Zeitgeber times 0, 4, 8, 16 and 20; for one *Por* sample extraction failed; RNA extraction and reverse transcription as above). qPCR was performed with Power SYBR Green PCR Master Mix on a StepOnePlus Real Time System (both Applied Biosystems). Fold-changes were calculated according to [72] in a custom excel sheet. The assumption of equal variance was violated for the RD comparison (F-Test) and the assumption of normal distribution was violated for the data of RA and RC in the *Por* strain (Shapiro-Wilk normality test), possibly reflecting circadian effects in the samples from different times of day. Thus, expression differences were assessed for significance in a two-tailed Wilcoxon Rank Sum Test (*wilcox.test* in R[66]). Holm correction[73] was used for multiple testing (default in *p.adjust* function of R).

CaMKII.1 minigenes

PCR fragments containing the CaMKII.1 linker region (exons 10 to 15) were amplified from genomic *Por* or *Jean* DNA respectively with primers CaMKII-Sc61-F-344112 and CaMKII-Sc61-R-351298 (Table S15), cloned with the CloneJET PCR Cloning Kit (Thermo Scientific), transferred into the pcDNA3.1+ vector using *NotI* and *XbaI* (Thermo Scientific). These constructs were transfected into Drosophila S2R+ cells and RNA was prepared 48h post transfection. After DNAse digestion, isoform expression was analyzed by radioactive, splicing-sensitive RT-PCR (primers in Table S17) and Phosphorimager quantification as described[74]. Identity of isoforms is based on size and sequencing of PCR products. To test for

reproducibility, there were seven biological replicates (raw data in Table S18). As the assumptions of equal variance (F-Test) and normal distribution of data (Shapiro-Wilk normality test) were not violated, the significance of expression differences was assessed in unpaired, two-sided two-sample t-tests. Holm correction[73] was used for multiple testing (default in *p.adjust* function of R). S2R+ cells were obtained from the lab of S. Sigrist, regularly authenticated by morphology and routinely tested for absence of mycoplasma contamination. The entire experiment was reproduced several months later with three biological replicates (raw data in Table S18).

S2 cell luciferase assay

Firefly luciferase is driven from a *period* 3x69 promoter under control of the CLOCK and CYCLE protein[19,21]. The *Drosophila pAc-clk* construct was obtained from F. Rouyer, *pCopia-Renilla luciferase* and *per3x69-luc* reporter constructs from M. Rosbash, a [Ca^{2+}] independent mouse *CaMKII* (T286D) was provided by M. Mayford. The CaMKII inhibitor KN-93 was purchased from Abcam (#ab120980).

C. marinus Cyc, *C. marinus Clk* and *C. marinus CaMKII.1-RD* were cloned into the pAc5.1/V5-His A plasmid (Invitrogen) with stop codons before the tag. The Q5® Site-Directed Mutagenesis Kit (NEB) was used to make kinase dead and [Ca^{2+}] independent versions of *C. marinus CaMKII.1-RD* (primers see Table S17).

Drosophila S2 cells (Invitrogen) were cultured at 25° C in Schneider's *Drosophila* medium (Lonza) supplemented with FBS (10%, heat-inactivated, penicillin (100 U/ml), streptomycin (100 µg/ml) and 2 mM L-glutamine; Sigma). Cells were seeded into 24 well plates (800,000 cells/well) and transfected with Effectene transfection reagent (Qiagen) according to the manufacturer's instructions.

Experiment with mouse [Ca^{2+}] independent CaMKII: 25ng *pCopia-Renilla*, 10ng *per3x69-luc*, 0.5ng *Drosophila pAc-clk, 200ng* mouse *pAc-CaMKII-T286D.* Experiment with CaMKII inhibitor KN-93: 25ng *pCopia-Renilla*, 10ng *per3x69-luc*, *Drosophila* 0.5ng *pAc-clk*, various amounts of KN-93. Experiment with *C. marinus* genes: 25ng *pCopia-Renilla*, 10ng *per3x69-luc*, 100ng *C. marinus pAc-cyc*, 100ng *C. marinus pAc-clk,* 200ng *C. marinus CaMKII.1-RD-K42R* or 200ng *C. marinus CaMKII.1-RD-T286D.* In all experiments, the transfection mix was filled up to a total of 435ng DNA with empty pAc5.1/V5-His A vector per well. After 48 hours, cells were washed with PBS and lysed with Passive Lysis Buffer (Promega). Luciferase activities were determined on a Synergy H1 plate reader (Biotek) using a Dual-Luciferase Reporter Assay System (Promega). For each biological replicate three independent cell lysates were measured and their mean value determined. Firefly luciferase activity was normalized to Renilla luciferase activity and values were normalized to controls transfected with *Drosophila pAc-clk* or *C. marinus pAc-clk* and *C. marinus pAc-cyc*, respectively. S2 cells (Invitrogen/Life Technologies, Cat.no. R690-07) were regularly authenticated by morphology and routinely tested for absence of mycoplasma contamination (Lonza MycoAlert). Sample size was chosen to test for reproducibility.

Circadian free-run experiments

For circadian free-run experiments, culture boxes of the *Por*, *He* and *Jean* strains were transferred from standard LD (16:8) to constant dim light (LL; about 100 lux). Emerging adults were collected in 1-hour intervals by a custom made *C. marinus* fraction collector (similar to [75]) and counted once a day. Because collection was automated, the experimenter had no influence on the results and blinding was not necessary. As the circalunar clock restricts adult emergence to

few days, the circadian emergence rhythm can only be assessed over few days. Several culture boxes were transferred to LL at different time points. The resulting emergence data were combined for each strain using the switch to LL as a common reference point. We used the maximum number of available individuals. Free-running period was calculated as the mean interval between subsequent emergence peaks, weighing each peak by the number of individuals.

Extended Data Figure 1 The biology of *Clunio marinus*

(a) *C. marinus* is restricted to rocky shores (black lines), the localities differing in tidal regime (adapted from[67]). **(b, c)** Local strains show corresponding genetic adaptations in their circadian (b;[67]) and circalunar rhythms (c; He[1], Jean[5]). Timing was measured in the laboratory under artificial moonlight (arrows in c) in a 30-day cycle and LD 12:12 (He, Por, Jean, Vigo) or 16:8 (Ber). Seasonal differences in daily illumination duration do not affect circadian emergence peaks[1,76]. Historically, for *C. marinus* "Zeitgeber time 0" is defined as the middle of dark phase.

Extended Data Figure 2 The reconstructed chromosomes of *C. marinus* based on the genetic linkage map

Left map: male informative markers. Right map: female informative markers. See Fig. 1a legend for further details.

Extended Data Figure 3 *C. marinus* genome characterization

(a) Representative genomic region with densely packed gene models (superscaffold1, from 535kb to 565kb). Gene models are given in blue on turquois background. Gene predictions (SNAP) are purple. Transcript evidence is yellow. **(b)** Phylogenetic relationships of *C. marinus to* other Diptera (according to [77]). **(c)** Genetic diversity (θ; red) and linkage disequilibrium (r^2; blue) of the *Jean* strain plotted for the three *C. marinus* linkage groups, revealing characteristic signatures of telomeres and centromeres. **(d-f)** Synteny comparisons among the genomes of *C. marinus*, *A. gambiae* and *D. melanogaster* based on 5,388 1:1:1 orthologs.

Extended Data Figure 4 Synteny analyses of *C. marinus* chromosome arms

928 **(a)** Gene content of the *C. marinus* chromosome arms relative to the
929 chromosome arms of *D. melanogaster* (black bars) and *A. gambiae* (grey bars).
930 The very small chromosome 4 of *D. melanogaster* is neglected. Chromosome
931 arms of *D. melanogaster* and *A. gambiae* are paired according to their published
932 homology (Zdobnov et al. 2002). For four of the chromosome arms of *C. marinus*
933 the homologous arms in *D. melanogaster* and *A. gambiae* are identified (grey
934 shading). For comparison, the conservation of the identified *D. melanogaster* and
935 *A. gambiae* homologs to each other is given by plotting the gene content of the
936 homologous *D. melanogaster* chromosome arm relative to the different
937 chromosome arms of *A. gambiae* (white bars). The numbers of orthologous
938 genes considered in each comparison are given above the bars. For chromosome
939 arm 2R of *C. marinus* the homologies are unclear. Possibly, chromosome arm 2R
940 of *C. marinus* has undergone so many re-arrangements with other chromosome
941 arms that it is no longer recognizable, which is consistent with complex
942 polymorphic re-arrangements in this chromosome arm of *C. marinus* (see
943 Supplementary Note 3). **(b)** Microsynteny is analyzed relative to *D. melanogaster*
944 and *A. gambiae*, based on 5.388 1:1:1 orthologs. The fraction of genes in
945 conserved microsynteny blocks is calculated and distributed along the
946 phylogenetic tree. **(c, d)** A simulation was used to estimate how many
947 chromosomal re-arrangements are required to produce the observed degree of
948 microsyntenic conservation (for details see Supplementary Note 2).

949 **Extended Data Figure 5 Population genomic analysis of QTLs C1/L1 and**
950 **C2 and genome-wide analysis of locations and putative effects of SNPs and**
951 **indels**

952 **(a, b)** Population genomic analysis of QTLs C1/L1 and C2.

953 Panels 1-3: *Por* vs. *Jean* strains (blue vs. red in panel 2,3). *Panel 1*: Genetic
954 differentiation (red dots: SNPs with $F_{ST} \geq 0.8$; grey dots: $F_{ST} < 0.8$; back line:
955 average F_{ST} in 5-kb sliding windows). *Panel 2:* Genetic diversity (θ) in 20-kb (thin
956 line) and 200-kb (thick line) windows. *Panel 3:* Linkage disequilibrium (r^2) for
957 SNP pairs from 0-600 bp apart in 100-kb windows (step size: 5kb). *Panel 4:*
958 Correlation Score (CS; 0 to 5) for genetic differentiation with circadian timing
959 (top), circalunar timing (middle) and geographic distance (bottom) for five
960 European *C. marinus* strains (*Vigo*, *Jean*, *Por*, *He*, *Ber*). Bottom numbers: scaffold
961 IDs. See also Fig. 1. **(c,d)** Locations and putative effects of SNPs (c) and indels (d)
962 with respect to the annotated gene models. The fractions of SNPs or indels in
963 each category are compared for all SNPs and indels (black bars) vs. differentiated
964 SNPs and indels (FST ≥ 0.8 for *Por* vs. *Jean* strain; grey bars). Absolute numbers
965 are given above the bars. In gene models with several splice forms, SNPs and
966 indels can have different effects, e.g. "CDS: non-synonymous" for one splice form
967 and "intronic" for another splice form. Therefore, the sum across locations is
968 slightly larger than the actual numbers of SNPs and indels. "Codon changes" are
969 all codon insertions or deletions that do not result in frame shifts beyond the
970 actual insertion/deletion site. CDS = coding sequence; syn. = synonymous; non-
971 syn. = non-synonymous; UTR = untranslated region.

972 **Extended Data Figure 6 CaMKII regulates CLK/CYC transcriptional**
973 **activity and exhibits strain specific splice variants**

974 **(a)** Quantification of luciferase activity under the control of an artificial 3x69 E-
975 box containing enhancer in S2 cells. Increasing amounts of the CaMKII inhibitor
976 KN-93 decrease luciferase activity in a concentration-dependent manner,
977 evidencing that endogenous CaMKII activity regulates the transcriptional activity

of the transfected CLOCK/CYCLE. **(b)** Without co-transfection of Drosophila *clock*, there is no detectable luciferase activity. The constitutively active form of CaMKII (mouse T286D) increases luciferase activity (normalised to CLOCK+; means; error bars: S.E.M.; biological replicates: n=4). **(c)** RNA sequencing reads mapped to the *CaMKII.1* genomic locus. Arrows: major differences between the strains. **(d)** Relative expression levels of the four major CaMKII.1 transcripts (RA to RD) and the minor variant RO in the *Por* and *Jean* strains of *C. marinus*, as measured by qPCR (mean values; error bars: S.E.M.; two-sided Wilcoxon rank sum test; *** $p<0.0005$; * $p<0.05$; "ns" is not significant; Holm correction for multiple testing; biological replicates: *Por* n=9, *Jean* n=10; except for RO: *Por* n=3, *Jean* n=8). RO was not detectable in six additional biological replicates of the *Por* strain, suggesting that the expression differences are even greater than currently estimated. Fig. 3a shows the same data, normalized to the respective *Por* strain variants.

Extended Data Figure 7 A differentiated 125bp insertion in the CaMKII locus

(a) Alignment of the part of the CaMKII locus of the *Por* and *Jean* strains that carries a 125bp insertion in the *Por* strain. **(b)** Pool-Seq reads (>150x coverage) of this position for *Por* and *Jean*, as shown in the Integrated Genome Viewer (IGV). The reference does not have the 125bp insertion. At the position marked by the red box, the *Jean* strain has a 4bp polymorphic indel (ATAC; frequently misaligned due to a SNP 8bp downstream), whereas the *Por* strain has the 125bp insertion (but not the 4bp ATAC insertion). In *Jean* basically all reads span the indel, suggesting that if the 125bp insertion is present in *Jean* at all, its frequency

is very low. In contrast, in *Por* all reads but one end at this position, suggesting the frequency of the 125bp insertion in *Por* is 154 of 155 reads or >0.99.

Extended Data Figure 8 Model of circadian timing adaptation via sequence differences in the *CaMKII.1* locus

Exon coloration as in Figure 4b. The arrows with question marks indicate possible pathways that alone or in combination could mediate the effect of CaMKII.1 on timing. Dotted lines: indirect effects.

Extended Data Figure 9 Analyses overview

(a) Overview of the genome assembly process. **(b)** Overview of the population genomic analyses.

Extended Data Figure 10 Arrangement of the mitochondrial genome (a) and of the histone gene cluster (b) of *C. marinus*.

Protein coding genes are given in yellow, tRNAs and rRNAs in red.

References for Online Methods and Extended Data Figures

46 Reineke, A., Karlovsky, P. & Zebitz, C. P. W. Preparation and purification of DNA from insects for AFLP analysis. *Insect Molecular Biology* **7**, 95-99 (1998).

47 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).

48 Aronesty, E. Command-line tools for processing biological sequencing data. http://code.google.com/p/ea-utils (2011).

49 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791, doi:10.1093/bioinformatics/btt468 (2013).

50 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

51 Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).

52 Gladman, S. & Seemann, T. VelvetOptimiser. http://bioinformatics.net.au/software.velvetoptimiser.shtml (accessed 2014).

53 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579 (2011).

54 Barton, M. D. & Barton, H. A. Scaffolder - software for manual genome scaffolding. *Source code for biology and medicine* **7**, 4-4, doi:10.1186/1751-0473-7-4 (2012).

55 Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biology* **13**, R56, doi:10.1186/gb-2012-13-6-r56 (2012).

56 Andrews, S. *FastQC A Quality Control tool for High Throughput Sequence Data*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 2015).

57 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652, doi:10.1038/nbt.1883 (2011).

58 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, doi:10.1186/1471-2105-12-491 (2011).

59 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, II215-II225, doi:10.1093/bioinformatics/btg1080 (2003).

60 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, doi:10.1186/1471-2105-5-59 (2004).

61 Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

62 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, I333-I339, doi:10.1093/bioinformatics/bts378 (2012).

63 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

64 McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

65 Kofler, R., Pandey, R. V. & Schlötterer, C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435-3436, doi:10.1093/bioinformatics/btr589 (2011).

66 Crawley, M. J. *The R Book*. (John Wiley & Sons Ltd., 2007).

67 Kaiser, T. S., Neumann, D., Heckel, D. G. & Berendonk, T. U. Strong genetic differentiation and postglacial origin of populations in the marine midge *Clunio marinus* (Chironomidae, Diptera). *Molecular Ecology* **19**, 2845-2857, doi:10.1111/j.1365-294X.2010.04706.x (2010).

68 Kofler, R. *et al.* PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE* **6**, e15925, doi:10.1371/journal.pone.0015925 (2011).

69 Feder, A. F., Petrov, D. A. & Bergland, A. O. LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. *PLoS ONE* **7**, e48588, doi:10.1371/journal.pone.0048588 (2012).

70 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. *Fly* **6**, 80-92, doi:10.4161/fly.19695 (2012).

71 Szkiba, D., Kapun, M., von Haeseler, A. & Gallach, M. SNP2GO: Functional Analysis of Genome-Wide Association Studies. *Genetics* **197**, 285-289, doi:10.1534/genetics.113.160341 (2014).

72 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

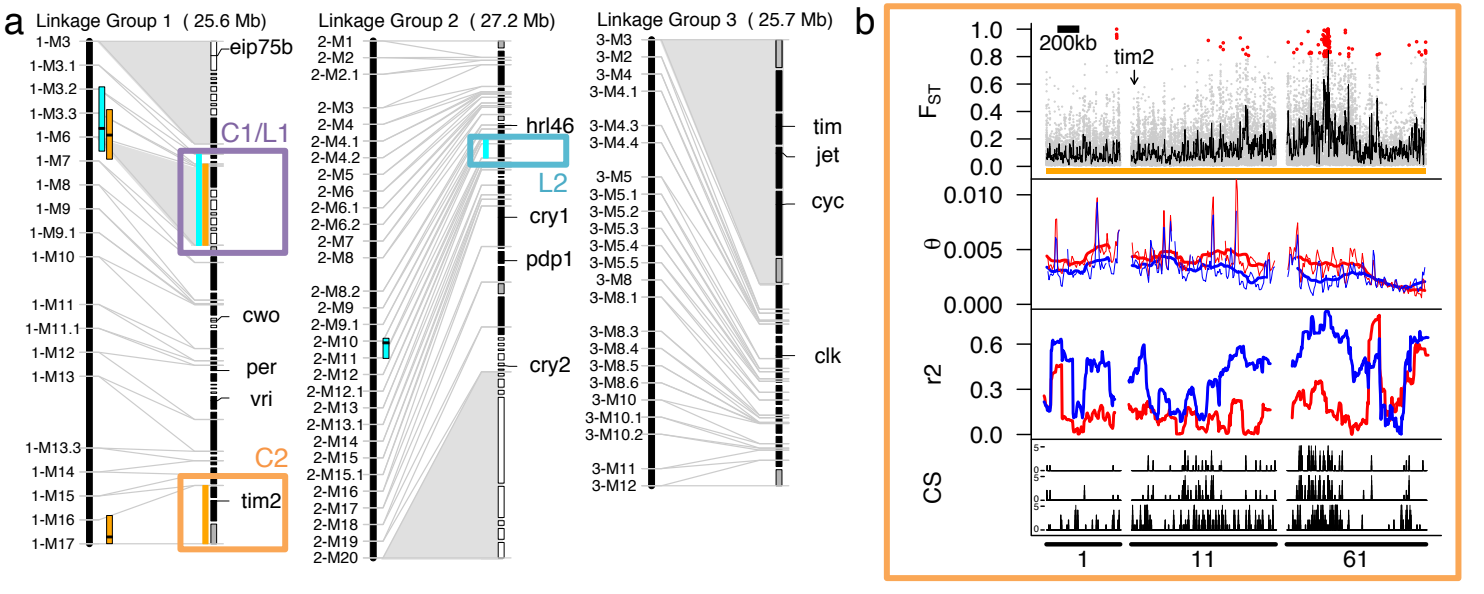
73 Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70 (1979).

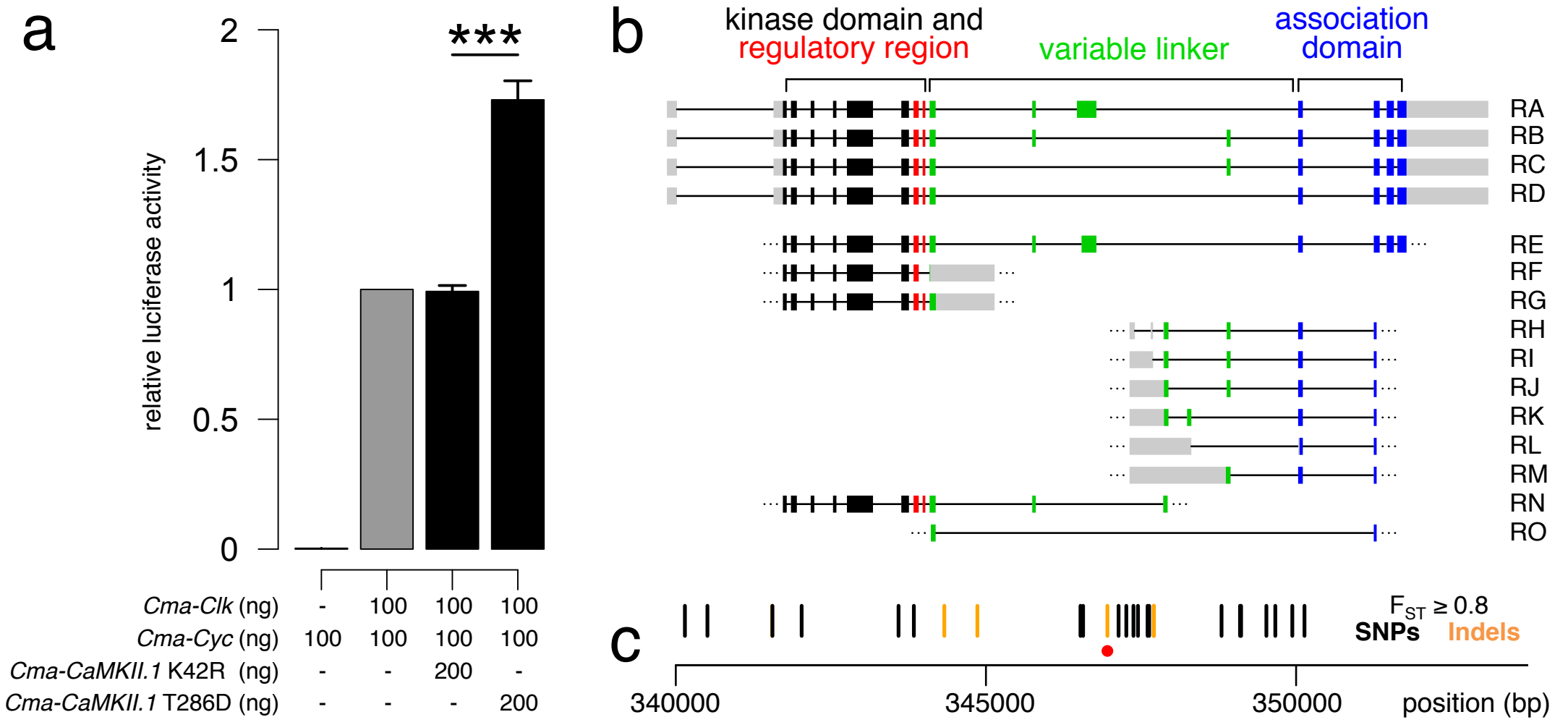
74 Preußner, M. *et al.* Rhythmic U2af26 Alternative Splicing Controls PERIOD1 Stability and the Circadian Clock in Mice. *Molecular Cell* **54**, 651-662, doi:http://dx.doi.org/10.1016/j.molcel.2014.04.015 (2014).

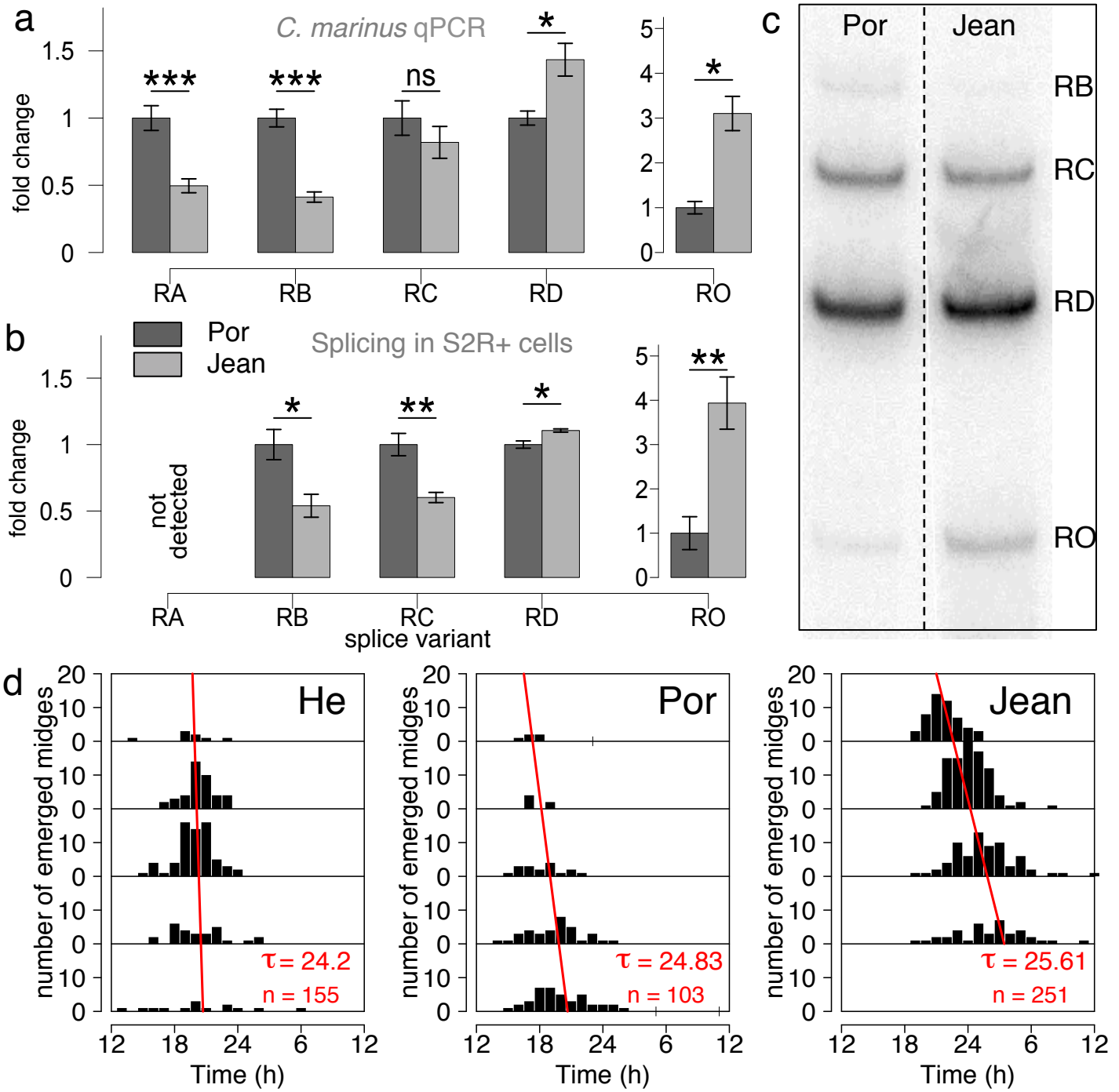
75 Honegger, H. W. An automatic device for the investigation of the rhythmic emergence pattern of Clunio marinus. *International Journal of Chronobiology* **4**, 217-221 (1977).

76 Heimbach, F. *Semilunare und diurnale Schlüpfrhythmen südenglischer und norwegischer Clunio-Populationen (Diptera, Chironomidae)* Dr thesis, Universität Köln, (1976).

77 Friedrich, M. & Tautz, D. Evolution and phylogeny of the diptera: A molecular phylogenetic analysis using 28S rDNA sequences. *Systematic Biology* **46**, 674-698, doi:10.2307/2413500 (1997).


a
Linkage Group 1 (25.6 Mb)
Linkage Group 2 (27.2 Mb)
Linkage Group 3 (25.7 Mb)
1-M3
1-M3.1
1-M3.2
1-M3.3
1-M6
1-M7
1-M8
1-M9
1-M9.1
1-M10
1-M11
1-M11.1
1-M12
1-M13
1-M13.3
1-M14
1-M15
1-M16
1-M17
eip75b
C1/L1
cwo
per
vri
C2
tim2
2-M1
2-M2
2-M2.1
2-M3
2-M4
2-M4.1
2-M4.2
2-M5
2-M6
2-M6.1
2-M6.2
2-M7
2-M8
2-M8.2
2-M9
2-M9.1
2-M10
2-M11
2-M12
2-M12.1
2-M13
2-M13.1
2-M14
2-M15
2-M15.1
2-M16
2-M17
2-M18
2-M19
2-M20
hrl46
L2
cry1
pdp1
cry2
3-M3
3-M2
3-M4
3-M4.1
3-M4.3
3-M4.4
3-M5
3-M5.1
3-M5.2
3-M5.3
3-M5.4
3-M5.5
3-M8
3-M8.1
3-M8.3
3-M8.4
3-M8.5
3-M8.6
3-M10
3-M10.1
3-M10.2
3-M11
3-M12
tim
jet
cyc
clk
b
200kb
tim2
F_{ST}
1.0
0.8
0.6
0.4
0.2
0.0
θ
0.010
0.005
0.000
r2
0.6
0.3
0.0
CS
1
11
61


a
***
relative luciferase activity
2
1.5
1
0.5
0
Cma-Clk (ng) - 100 100 100
Cma-Cyc (ng) 100 100 100 100
Cma-CaMKII.1 K42R (ng) - - 200 -
Cma-CaMKII.1 T286D (ng) - - - 200
b
kinase domain and regulatory region
variable linker
association domain
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
c
F_ST ≥ 0.8
SNPs
Indels
340000
345000
350000
position (bp)


a
C. marinus qPCR
fold change
***
***
ns
*
*
RA
RB
RC
RD
RO
b
Por
Jean
Splicing in S2R+ cells
fold change
not detected
*
**
*
**
RA
RB
RC
RD
RO
splice variant
c
Por
Jean
RB
RC
RD
RO
d
He
Por
Jean
number of emerged midges
τ = 24.2
n = 155
τ = 24.83
n = 103
τ = 25.61
n = 251
Time (h)